

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

DAILY NEWS, LP; CHICAGO TRIBUNE COMPANY, LLC; ORLANDO SENTINEL COMMUNICATIONS COMPANY, LLC; SUN-SENTINEL COMPANY, LLC; SAN JOSE MERCURY- NEWS, LLC; DP MEDIA NETWORK, LLC; ORB PUBLISHING, LLC; and NORTHWEST PUBLICATIONS, LLC,

Plaintiffs,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL, LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Case No. 1:24-cv-03285-SHS

**DEFENDANT MICROSOFT CORPORATION'S MEMORANDUM
IN SUPPORT OF PARTIAL MOTION TO DISMISS THE COMPLAINT**

TABLE OF CONTENTS

PRELIMINARY STATEMENT	1
ALLEGATIONS OF THE COMPLAINT.....	2
A. Large Language Models Are A Profound Advance In Artificial Intelligence And A Powerful Tool For Human Flourishing.....	2
B. Microsoft Collaborates With OpenAI To Bring GPT-Based Tools To The Public.	4
C. Eight Publishers Sue Microsoft and OpenAI.....	4
D. The Complaint Alleges No Real-World Instance of Output of Plaintiffs' Works or Impact to Plaintiffs' Rights.....	6
ARGUMENT	7
I. PLAINTIFFS FAIL TO STATE A CLAIM FOR CONTRIBUTORY INFRINGEMENT BASED ON USER OUTPUTS.	8
A. The Complaint Fails To Allege An Actionable Instance Of End-User Copyright Infringement.....	8
B. The Complaint Fails To Allege Microsoft's Knowledge Of Direct Infringement.	10
II. PLAINTIFFS FAIL TO STATE A DMCA § 1202 CLAIM.....	12
A. The Complaint Does Not Plausibly Allege Microsoft's Intent To Induce, Enable, Facilitate, Or Conceal Infringement By Removal Of CMI.....	13
B. The Complaint Does Not Plausibly Allege "Removal" Of CMI From Copies Of Identical Works.....	17
III. PLAINTIFFS' MISAPPROPRIATION CLAIM IS PREEMPTED BY THE COPYRIGHT ACT.....	20
IV. PLAINTIFFS' STATE LAW DILUTION CLAIM IS BARRED BY THE DORMANT COMMERCE CLAUSE.....	23
CONCLUSION.....	25

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>A'Lor Int'l, Ltd. v. Tappers Fine Jewelry, Inc.</i> , No. 12-cv-02215, 2012 WL 12921035 (C.D. Cal. Aug. 8, 2012).....	19
<i>ACLU v. Johnson</i> , 194 F.3d 1149.....	24
<i>Am. Libraries Ass'n. v. Pataki</i> , 969 F. Supp. 160 (S.D.N.Y. 1997).....	23
<i>Andersen v. Stability AI Ltd.</i> , No. 23-cv-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023).....	12, 17
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009)	7, 8, 14
<i>BMG Rights Mgmt. (US) LLC v. Cox Commc'ns, Inc.</i> , 881 F.3d 293 (4th Cir. 2018).....	10
<i>Ctr. for Democracy & Tech. v. Pappert</i> , 337 F. Supp. 2d 606 (E.D. Pa. 2004).....	24
<i>DBW Partners, LLC v. Mkt. Sec., LLC</i> , No. 22-cv-1333, 2023 WL 2610498 (D.D.C. Mar. 23, 2023)	23
<i>Design Basics, LLC v. WK Olson Architects, Inc.</i> , No. 17-cv-7432, 2019 WL 527535 (N.D. Ill. Feb. 11, 2019).....	18
<i>Doe 1 v. GitHub, Inc.</i> , No. 22-cv-6823, 2024 WL 235217 (N.D. Cal. Jan. 22, 2024)	12, 17, 18, 19
<i>Dolls Kill, Inc. v. Zoetop Bus. Co.</i> , No. 22-cv-01463, 2022 WL 16961477 (C.D. Cal. Aug. 25, 2022)	19
<i>Edgar v. MITE Corp.</i> , 457 U.S. 624 (1982)	23
<i>Falkner v. Gen. Motors LLC</i> , 393 F. Supp. 3d 927 (C.D. Cal. 2018).....	18
<i>Fashion Nova, LLC v. Blush Mark, Inc.</i> , No. 22-cv-6127, 2023 WL 4307646 (C.D. Cal. June 30, 2023).....	14
<i>Faulkner Press, L.L.C. v. Class Notes, L.L.C.</i> , 756 F. Supp. 2d 1352 (N.D. Fla. 2010)	18

<i>Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,</i> 499 U.S. 340 (1991)	9
<i>Fin. Info., Inc. v. Moody's Invs. Serv., Inc.,</i> 808 F.2d 204 (2d Cir. 1986)	22
<i>Fischer v. Forrest,</i> 286 F. Supp. 3d 590 (S.D.N.Y. 2018)	19
<i>Free Speech Sys., LLC v. Menzel,</i> 390 F. Supp. 3d 1162 (N.D. Cal. 2019).....	17
<i>Frost-Tsuji Architects v. Highway Inn, Inc.,</i> No. 13-cv-00496, 2015 WL 263556 (D. Haw. Jan. 21, 2015)	19
<i>Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc.,</i> 443 F.2d 1159 (2d Cir. 1971)	8, 10
<i>Greer v. Fox News Media,</i> No. 22-1970, 2023 WL 2671796 (2d Cir. Mar. 29, 2023)	23
<i>Healy v. Beer Inst. Inc.,</i> 491 U.S. 324 (1989)	23
<i>Kelly v. Arriba Soft Corp.,</i> 77 F. Supp. 2d 1116 (C.D. Cal. 1999).....	18
<i>Ludvarts, LLC v. AT&T Mobility, LLC,</i> 710 F.3d 1068 (9th Cir. 2013).....	10, 11
<i>Mango v. BuzzFeed, Inc.,</i> 970 F.3d 167 (2d Cir. 2020)	13, 15
<i>Matthew Bender & Co. v. West Publ'g Co.,</i> 158 F.3d 693 (2d Cir. 1998)	8, 9
<i>MGM Studios Inc. v. Grokster, Ltd.,</i> 545 U.S. 913 (2005)	12
<i>ML Genius Holdings LLC v. Google LLC,</i> No. 20-3113, 2022 WL 710744 (2d Cir. Mar. 10, 2022)	23
<i>Nat'l Pork Producers Council v. Ross,</i> 598 U.S. 356 (2023)	23
<i>National Basketball Ass'n v. Motorola, Inc.,</i> 105 F.3d 841 (2d Cir. 1997)	21, 22
<i>Publius v. Boyer-Vine,</i> 237 F. Supp. 3d 997 (E.D. Cal. 2017)	24

Roberts v. BroadwayHD LLC,
518 F. Supp. 3d 719 (S.D.N.Y. 2021)..... 14

Sony Music Ent. v. Cox Commc'ns, Inc.,
93 F.4th 222 (4th Cir. 2024)..... 10

Stevens v. Corelogic, Inc.,
899 F.3d 666 (9th Cir. 2018)..... 13, 16, 17

Tremblay v. OpenAI, Inc.,
No. 23-cv-03223, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024)..... 9, 12, 14, 17

Viacom Int'l, Inc. v. Youtube, Inc.,
676 F.3d 19 (2d Cir. 2012)..... 10

Statutes

15 U.S.C. § 1125(c), 5

Copyright Act, 17 U.S.C. §§ 101 *et seq.*

 § 101 17

 § 107 5, 15

 § 301(a) 20

 § 1202(b) 12, 17

 § 1202(b)(1) 6, 12

 § 1202(b)(3) 6, 12, 17

New York's Gen. Bus. Law § 360-l 6, 23, 24

Other Authorities

Fed. R. Civ. P. 12(b)(6)..... 9

S. Rep. No. 105-190 (1998) 12

Webster's Third New International Dictionary (2002) 17, 18

PRELIMINARY STATEMENT

Eight newspapers have filed this follow-on lawsuit challenging generative AI tools offered by Microsoft and OpenAI. As Plaintiffs acknowledge, their Complaint largely repeats the claims and allegations from the Complaint in *The New York Times v. Microsoft*, a related case pending before this Court. So, unsurprisingly, this Complaint contains the same deficiencies—including fatal defects in the contributory copyright infringement, DMCA, and state law misappropriation claims addressed in this motion to dismiss. But two aspects of the Complaint here are uniquely telling.

First, this Complaint omits The Times’ unsubstantiated prediction that Microsoft and OpenAI’s tools will somehow destroy independent journalism—evidently a talking point these Plaintiffs do not believe they can substantiate. And like The Times’ Complaint, this one elides the key technical and legal fact that will ultimately decide this case: Microsoft and OpenAI’s tools neither exploit the protected expression in the Plaintiffs’ digital content nor replace it—they extract and share elements of language, culture, ideas, and knowledge that belong to all of us. That is why The Times’ and Plaintiffs’ claims ultimately fail under the doctrine of fair use.

Second, much like The Times, Plaintiffs here *assert* that Defendants’ GPT-based products are capable of being prompted to output snippets of Plaintiffs’ protected works, yet point to no example generated under real-world conditions. What is so telling this time is that Plaintiffs filed their Complaint four months after The Times, and a month after the motions to dismiss in that case were fully briefed—despite every incentive and months of opportunity to locate some real-world substantiation for their claims, they came up with nothing. They therefore proceed on the basis of the same contrived prompts and rigged outputs as The Times did—examples that show at most what highly versatile LLMs are *capable* of doing, not what they have or will actually do when used by the public.

Plaintiffs' failures merit dismissal of the same three claims Microsoft moved to dismiss from The Times' Complaint. The contributory infringement claim against Microsoft based on purported end-user infringement fails because Plaintiffs have not alleged either a real-world instance of infringement of their works, nor Microsoft's knowledge of specific instances of infringement to which it could have knowingly contributed. Part I, *infra*. Plaintiffs' DMCA § 1202 claims fail because they can allege neither that removal of CMI from their works would ever lead to piracy of those works, nor that the GPT-based products would ever output the sorts of identical copies of works that are required under § 1202. Part II, *infra*. And Plaintiffs' attempt at a state-law misappropriation claim based on "hot news" is just a copyright claim in disguise, and therefore is preempted by the Copyright Act. Part III, *infra*. These claims should be dismissed.

Finally, Plaintiffs assert a state-law trademark dilution claim not presented in The Times' Complaint. By asserting this claim, Plaintiffs seek to deploy New York law to directly regulate transactions occurring wholly out of state, governing what content may appear on computer screens nationwide, in plain violation of the dormant Commerce Clause. Part IV, *infra*. This claim too should be dismissed.

ALLEGATIONS OF THE COMPLAINT

A. Large Language Models Are A Profound Advance In Artificial Intelligence And A Powerful Tool For Human Flourishing.

Over the past several years, scientists and engineers have made dramatic advances in the field of artificial intelligence. Perhaps the most astounding is the "large language model," or LLM. As relevant here, an LLM is a machine learning model that can process and produce natural language text. Compl. ¶¶ 58, 61, 73, 78.

At the heart of an LLM is a computer algorithm called a “transformer.” *See* Compl. ¶ 80. To create a model, engineers engage in a step called “training,” in which they feed large amounts of text through the transformer. Compl. ¶¶ 72, 74. As this happens, the LLM separates the text into “tokens”—words, punctuation, and so forth. *See* Compl. ¶ 84. Across the entire “training set” or “training corpus,” the model then discerns semantic patterns among tokens, which it “encode[s] … as numbers called ‘parameters.’” Compl. ¶¶ 74, 88. A highly sophisticated model trained on a large corpus might have trillions of tokens and parameters. *See, e.g.*, Compl. ¶ 88 (alleging that GPT-4 contains “13 trillion tokens” and “1.8 trillion parameters”). It is this extensive network of semantic connections that allows the model to emit natural-language text. Compl. ¶¶ 72-74.

LLMs are sometimes referred to as “general purpose” or “foundational” AI models because their language capabilities have many applications. “Once trained, LLMs may be provided with information specific to a use case or subject matter,” Compl. ¶ 78, enabling them to assist a human user with that use case or subject. That user directs the LLM’s performance through user-selected “prompts” to which the model responds. Compl. ¶¶ 58, 73.

As reflected in various sources incorporated by reference into the Complaint, Microsoft believes that LLMs “can so deeply absorb the nuances of language, grammar, knowledge, concepts, and context that [they] can excel at multiple tasks,” promising untold applications that will improve people’s lives. Jennifer Langston, *Microsoft announces new supercomputer; lays out vision for future AI work*, news.microsoft.com/source/features/ai/openai-azure-supercomputer (May 19, 2020) (cited at Compl. ¶ 67 n.11). Microsoft has also urged the need for a responsible approach to such powerful technology, routinely stressing the importance of “responsible AI and AI safety” from the highest levels of the company. Compl. ¶¶ 68, 91.

B. Microsoft Collaborates With OpenAI To Bring GPT-Based Tools To The Public.

OpenAI was “formed in December 2015 as a ‘non-profit artificial intelligence research company.’” Compl. ¶ 52. OpenAI has developed several versions of its foundational LLMs, built on its “Generative Pre-training Transformer” or GPT. Compl. ¶¶ 19, 55-56. “OpenAI became a household name upon the release of ChatGPT,” a “text-generating chatbot” powered by an underlying GPT model, that “given user-generated prompts, can mimic humanlike natural language responses.” Compl. ¶ 58. Since 2019, Microsoft has collaborated with OpenAI to bring GPT-based products to the public, providing technological infrastructure and cloud computing services. Compl. ¶¶ 63, 65, 67.

The Microsoft-OpenAI collaboration has helped bring the immense promise of LLMs to the public. Compl. ¶ 79. Microsoft offers “Azure OpenAI Service,” used by “18,000 organizations.” Compl. ¶ 30. It also offers “Bing Chat (now Copilot), a generative AI chatbot feature on its search engine powered by GPT-4.” Compl. ¶ 69. And “Microsoft and OpenAI unveiled ‘Browse with Bing,’ a plugin to ChatGPT that enabled it to access the latest content on the internet through the Microsoft Bing search engine.” Compl. ¶ 69.

C. Eight Publishers Sue Microsoft and OpenAI.

The newspapers that bring this lawsuit “publish[] digital and print products.” Compl. ¶¶ 21-28. Plaintiffs allege that “[m]illions of [their] Works were copied and ingested” in training OpenAI’s GPT models, such that pieces of these works are among the trillions of tokens and parameters from which predictive natural language responses are generated. Compl. ¶¶ 88, 90.

The Complaint alleges eight causes of action against Microsoft based on multiple legal theories. Compl. ¶¶ 190-254. The first count of the Complaint alleges that Microsoft engaged in direct copyright infringement based on theories involving “training” of OpenAI’s models,

Compl. ¶ 193; “storing, processing, and reproducing the GPT models … on Microsoft’s supercomputing platform,” Compl. ¶ 195; and “disseminating generative output” to users, Compl. ¶ 196. The second two counts allege vicarious and contributory infringement against Microsoft for its role in the direct infringement by OpenAI alleged in Count I. Compl. ¶¶ 202-06 (vicarious infringement); Compl. ¶¶ 207-10 (contributory infringement). These three claims are not at issue in this motion. These claims will ultimately turn, among other things, on a fair-use defense under 17 U.S.C. § 107 not suited to the pleading stage. The Complaint’s seventh cause of action—a trademark tarnishment theory suggesting that “lower-quality and inaccurate” output could “dilute[] the quality of the Diluted Trademarks”—is also not at issue in this motion. Compl. ¶¶ 233-49; *see* 15 U.S.C. § 1125(c).

At issue here are Counts IV, V, VI, and VIII. The fourth count alleges an alternative theory of contributory infringement against Microsoft and OpenAI. Compl. ¶¶ 211-14. This claim anticipates another defense to one of the direct infringement claims—that it is the *user*, not Microsoft or OpenAI, who is the direct infringer of a “user-generated prompt[],” Compl. ¶ 58, resulting in an infringing output. The Complaint alleges that “to the extent an end-user may be liable as a direct infringer based on output of the GPT-based products, Defendants materially contributed to and directly assisted with the direct infringement perpetrated by end-users.” Compl. ¶ 212. It claims that “Defendants had either actual knowledge or constructive knowledge” of infringement because they “have recognized … that GPT-based products *are capable* of distributing unlicensed copies of copyrighted works.” Compl. ¶ 214 (emphasis added).

The Complaint’s fifth cause of action alleges violation of § 1202 of the DMCA, which protects against the removal of “copyright management information,” or CMI, from copies of

works. The Complaint alleges that Microsoft and OpenAI “removed the Publishers’ copyright-management information” in “building the training datasets,” “generating synthetic search results,” and “generating outputs from the GPT models.” Compl. ¶¶ 218-21 (invoking 17 U.S.C. § 1202(b)(1)). It also alleges that Defendants “distribut[ed]” works knowing that CMI was removed. Compl. ¶ 223 (invoking 17 U.S.C. § 1202(b)(3)).

The Complaint’s sixth cause of action alleges common law unfair competition by misappropriation. Plaintiffs allege that “[b]y offering GenAI content that is the same as or similar to content published by the Publishers, Defendants’ GenAI products directly compete with the Publishers’ content.” Compl. ¶ 228. The Complaint alleges that Defendants misappropriate “time-sensitive breaking news,” Compl. ¶ 227, and that this “substantially threatens the quality of the Publishers’ content and disincentives the Publishers to produce their content.” Compl. ¶ 232.

The Complaint’s eighth cause of action alleges dilution of business reputation under New York’s General Business Law § 360-1, largely replicating the separate claim for trademark dilution under the Lanham Act. Compl. ¶¶ 250-54. Plaintiffs allege that they possess “[t]rademarks [with] a distinctive quality ... since long before Defendants adopted and began” allegedly “us[ing] ... the ... [t]rademarks.” Compl. ¶ 251. The Complaint alleges that “Defendants are aware that their GPT-based products produce inaccurate content that is falsely attributed to” Plaintiffs, and “profit commercially from creating and attributing inaccurate content to” Plaintiffs. Compl. ¶ 253.

D. The Complaint Alleges No Real-World Instance of Output of Plaintiffs’ Works or Impact to Plaintiffs’ Rights.

The Complaint is filled with attempts to establish that GPT-based products “will output near-verbatim copies of significant portions of Publishers’ Works when prompted to do so.”

Compl. ¶ 96; *see* Compl. ¶¶ 97-113. But it pleads no specific instance of infringement of Plaintiffs' works in the real world as a result of the development, offering, or use of GPT-based products. Plaintiffs say that "given the right prompt" these products "will repeat large portions" of Plaintiffs' works, Compl. ¶ 77, offering various examples that compare the text of an article to model output. But Plaintiffs bury "the right prompt" in an exhibit, which admits that the prompts each "comprise[] the beginning of an article from the Publisher"—often several verbatim *paragraphs* of the work the prompt is designed to elicit. Compl. Ex. J at 1. The Complaint does not explain why any real person, already in possession of an article by Plaintiffs, would ever want to feed a part of that prompt into a GPT-based product to generate the rest of the article. As mentioned, that person already has the entire article. The Complaint certainly does not allege that any real person has done this. This is a manufactured story.

Similarly unsubstantiated are Plaintiffs' various warnings of harm—from "threat[s] [to] the quality of the Publishers' content," Compl. ¶ 232, to hindering their "ability to attract and retain paying subscribers," Compl. ¶ 188, to "compet[ing] with the Publishers' content for traffic," Compl. ¶ 229, to "tarnishing the newspapers' reputations" "by falsely attributing inaccurate reporting to the Publishers' newspapers," Compl. ¶ 7. All of this turns out to be mere speculation about what Plaintiffs apparently fear *might* happen. Indeed, the Complaint does not allege a single real-world example of any occurrence that could even begin to cause such harm, let alone allege that any such harms have materialized.

ARGUMENT

To survive a motion to dismiss under Rule 12(b)(6), "a complaint must contain sufficient factual matter, accepted as true, to 'state a claim to relief that is plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (internal quotation marks omitted). If a claim lacks a cognizable theory, lacks sufficient facts to plausibly support that theory, or advances a theory that

is foreclosed as a matter of law, dismissal is appropriate. “Threadbare recitals of the elements of a cause of action, supported by mere conclusory statements,” are insufficient. *Id.*

I. PLAINTIFFS FAIL TO STATE A CLAIM FOR CONTRIBUTORY INFRINGEMENT BASED ON USER OUTPUTS.

Plaintiffs fail to state a claim that Microsoft has contributed to end-user copyright infringement. *See* Compl. ¶¶ 211-14 (Count IV). A contributory infringer is “one who, with knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another.” *Gershwin Publ’g Corp. v. Columbia Artists Mgmt., Inc.*, 443 F.2d 1159, 1162 (2d Cir. 1971). The Complaint does not allege that Microsoft encouraged end-users to make infringing use of any GPT-based product, acted in a way that would promote such infringing use, or did anything to induce such infringing use. It advances only a “material[] contribut[ion]” theory. Compl. ¶ 212. The notion is that by offering end-users products that it allegedly knows “are capable of distributing unlicensed copies” or “derivative[]” works, Microsoft is a contributory infringer any time a user makes use of that capability. Compl. ¶ 214.

This theory fails for two reasons. First, Plaintiffs fail to allege any instance of direct infringement by a user, a necessary prerequisite to a claim for secondary liability. Part I.A, *infra*. Second, Plaintiffs cannot allege that Microsoft knowingly participated in any specific infringement. Part I.B, *infra*.

A. The Complaint Fails To Allege An Actionable Instance Of End-User Copyright Infringement.

“Contributory infringement necessarily must follow a finding of direct or primary infringement.” *Matthew Bender & Co. v. West Publ’g Co.*, 158 F.3d 693, 706 (2d Cir. 1998) (quoting *Cable/Home Communication Corp. v. Network Prods., Inc.*, 902 F.2d 829, 845 (11th Cir.1990)). To allege direct infringement, a plaintiff must plead (1) its “ownership of a valid copyright,” and (2) a primary infringer’s “copying of constituent elements of the work that are

original”—*i.e.*, infringement of one of the rights protected under 17 U.S.C. § 106. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991).

In *Matthew Bender*, West Publishing Company, the publisher of the “National Reporter System,” sued the publisher of electronic databases that collected judicial opinions. 158 F.3d at 697. West alleged that a user of these databases could operate them to “view (and print) judicial opinions in the same order in which they are printed in a West volume,” thus recreating West’s reporters. *Id.* On this basis, it sought to hold the “database manufacturer … liable as a contributory infringer … for creating a product that assists a user to infringe a copyright directly.” *Id.* at 706. The Second Circuit rejected the claim. Though West “hypothesized that users … will retrieve and print cases in the order in which they appear in West’s case reporters,” it “failed to identify any primary infringer, other than [its own] counsel.” *Id.*

Plaintiffs have the same problem here. The Complaint “hypothesize[s]” that someone *could* find a way to prompt the GPT-based products to yield output that is similar to one of Plaintiffs’ works. It contains examples of prompts that yield small amounts of similar text—prompts like multi-paragraph passages of an article, Compl. Ex. J, or cumbersome, direct requests for specific paragraphs of a specific piece. Compl. ¶¶ 120-38. But this is purely use of the tools by Plaintiffs or individuals directed by them with instructions to manufacture infringement. Plaintiffs do not allege that any actual user of a GPT-based product has or would enter such prompts to elicit Plaintiffs’ works. The bare theoretical possibility that someone somewhere might engage in the same acrobatics Plaintiffs did here is not enough to plausibly allege direct infringement. This warrants dismissal of the claim. *See Tremblay v. OpenAI, Inc.*, No. 23-cv-03223, 2024 WL 557720, at *2-3 (N.D. Cal. Feb. 12, 2024) (dismissing vicarious

infringement claim based on ChatGPT outputs for failure to adequately allege direct infringement).

B. The Complaint Fails To Allege Microsoft’s Knowledge Of Direct Infringement.

The Complaint also fails to allege “knowledge of the infringing activity.” *Gershwin*, 443 F.2d at 1162. As *Gershwin* recognized, contributory liability is based on “the common law doctrine that one who *knowingly participates or furthers* a tortious act is jointly and severally liable with the prime tortfeasor.” *Id.* (emphasis added). A plaintiff must therefore allege more than mere “[g]eneral[] know[ledge] of infringement,” especially where the defendant offers a product or service with noninfringing uses. *BMG Rights Mgmt. (US) LLC v. Cox Commc’ns, Inc.*, 881 F.3d 293, 311 (4th Cir. 2018). “Selling a product with both lawful and unlawful uses suggests an intent to cause infringement only if the seller knows of *specific* instances of infringement.” *Id.* at 311; *accord Ludvarts, LLC v. AT&T Mobility, LLC*, 710 F.3d 1068, 1072-73 (9th Cir. 2013) (demanding knowledge of “specific acts of infringement”); *cf. Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 32 (2d Cir. 2012) (knowledge standard under safe harbor provision of DMCA, 17 U.S.C. § 512(c), requires “awareness of facts or circumstances that indicate specific and identifiable instances of infringement”).

The bedrock requirement of knowing participation also demands that the defendant have “actual knowledge” of (or be willfully blind to) specific acts of infringement. 710 F.3d 1068, 1072-73. It is not enough that a defendant “should have known of … infringing activity.” *BMG*, 881 F.3d at 310. And as a logical matter, the defendant must have knowledge of the specific act before it takes place—that is, “predictive” knowledge—because one cannot knowingly participate in an act learned of only after it is completed. *Sony Music Ent. v. Cox Commc’ns, Inc.*, 93 F.4th 222, 234 (4th Cir. 2024).

The Complaint does not come close to meeting this high bar. It alleges that Defendants knew or had reason to know of “the direct infringement by end-users because … Defendants undertake extensive efforts in developing, testing, and troubleshooting their LLM models and GPT-based products.” Compl. ¶ 214. It also says that “the capability of their GPT-based models to produce infringing output has been the subject of public conversation,” that “Defendants are aware that at least some users use their GPT-based products for the purpose of accessing copyrighted works,” and that “Defendants have publicly recognized and admitted that their GPT-based products are capable of distributing unlicensed copies of copyrighted works.” Compl. ¶ 214. But these formulaic allegations do not identify a “specific act[] of infringement,” *Ludvarts*, 710 F.3d at 1072-73, let alone allege Microsoft’s actual knowledge of one before it takes place such that one could infer Microsoft’s knowing participation.

The same problem sinks Plaintiffs’ vague allegation that “Defendants programmed their systems to flag infringing outputs and prompts seeking infringing output.” Compl. ¶ 214. It is unclear what Plaintiffs mean by any of this—what it means to “flag” outputs or prompts, or what Microsoft allegedly did upon such flagging. But whatever Plaintiffs intend, they do not allege that Microsoft, via these alleged systems, knew of any specific act of infringement before or while it was happening, and then knowingly participated in that act. If anything, all this allegation appears to do is acknowledge that Microsoft is taking precautionary measures to ensure infringing outputs do not happen in the first place. This vague, bare insinuation therefore does not save Plaintiffs’ claim.

At most, Plaintiffs’ allegations establish Microsoft’s awareness that someone *could* use a GPT-based product to infringe. The same was true of the VCR—as it is of word processors, hard drives, social media feeds, internet connections, and so forth. Since *Sony Corp. of Am. v.*

Universal City Studios, the law has barred liability based on “the equivocal conduct of selling an item with substantial lawful as well as unlawful uses,” “limit[ing] liability to instances of more acute fault than the mere understanding that some of one’s products will be misused.” *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 932-33 (2005).

II. PLAINTIFFS FAIL TO STATE A DMCA § 1202 CLAIM.

Plaintiffs also fail to state a claim under either § 1202(b)(1) or (b)(3) of the DMCA. *See* Compl. ¶¶ 215-225 (Count V). Section 1202 is one of two adjoining provisions that aims to thwart digital piracy. S. Rep. No. 105-190 at 11 n.18 (1998) (DMCA aimed to “discourage piracy”). It protects the “[i]ntegrity of copyright management information”—defined as “information conveyed in connection with copies … of a work”—on the rationale that removal or falsification of that information aids the spread of pirated copies. *Id.* at 34, 89. To state a claim under § 1202(b)(1), a plaintiff must allege that the defendant “intentionally remov[ed] or alter[ed]” CMI from a copy of a work. 17 U.S.C. § 1202(b)(1). To state a claim under § 1202(b)(3), a plaintiff must plausibly allege that the defendant “distribute[d]” copies of works “knowing that [CMI] has been removed or altered without authority of the copyright owner[.]” And for both claims, a plaintiff must further establish that the defendant removed or distributed CMI “knowing, or … having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this title.” 17 U.S.C. § 1202(b).

Over the past year and a half, courts have repeatedly rejected challenges to generative AI tools based on § 1202.¹ These decisions establish that neither the highly technical process of training an LLM nor the generated outputs of that LLM implicate the sort of piracy concerns

¹ See *Tremblay*, 2024 WL 557720, at *4-5 (challenge to ChatGPT); *Doe I v. GitHub, Inc.*, No. 22-cv-6823, 2024 WL 235217, at *8-9 (N.D. Cal. Jan. 22, 2024) (challenge to GPT-based “Codex” model); *Andersen v. Stability AI Ltd.*, No. 23-cv-00201, 2023 WL 7132064, at *10-11 (N.D. Cal. Oct. 30, 2023).

animating § 1202. Section 1202 claims are fundamentally inapt when applied to a machine that generates brand new language in the form of conversational excerpts or snippets of information.

The claims here fail for the same two reasons the others have. First, Plaintiffs cannot allege Microsoft’s culpable awareness that removal of CMI will “induce, enable, facilitate, or conceal an infringement” of their works. Part II.A., *infra*. And second, they cannot allege CMI was removed from identical copies of their works, as required. Part II.B., *infra*.

A. The Complaint Does Not Plausibly Allege Microsoft’s Intent To Induce, Enable, Facilitate, Or Conceal Infringement By Removal Of CMI.

Because it is an anti-piracy statute, § 1202 is careful not to impose liability on a defendant whose removal of CMI has only the incidental effect of aiding or concealing copyright infringement. It enacts this limitation through a stringent “double-scienter requirement.” *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020). The first half of the requirement demands that the defendant’s conduct—removal of CMI, alteration of CMI, and so forth—be intentional (for subsection (b)(1)) or knowing (for subsection (b)(3)). *See id.* The second half, at issue here, requires that the “defendant know or have reason to know” that its conduct “will induce, enable, facilitate, or conceal an infringement.” *Id.*

Though this latter requirement is phrased as a scienter standard, courts have recognized that it carries an objective component in cases involving technology that may incidentally remove CMI. Unless a plaintiff can point to direct facts revealing the defendant’s subjective belief that removal of CMI will facilitate infringement, it must plausibly allege that, as a result of technology’s removal of CMI, infringement is *objectively* “likely.” *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675 (9th Cir. 2018). Otherwise there is no basis for inferring that a defendant “was aware or had reasonable grounds to be aware” that “the probable future impact of its actions” would be infringement. *Id.* at 674.

The Complaint fails to plausibly allege this likelihood. It alleges (mostly “upon information and belief”) that “Defendants” removed CMI “in building the training datasets,” Compl. ¶ 218, “through generating synthetic search results,” Compl. ¶ 219, and “in generating outputs from the GPT models,” Compl. ¶ 220. And it then advances only the conclusory assertion that the Defendants’ conduct “has been done knowingly and with the intent to induce, enable, facilitate, or conceal Defendants’ or end-users’ infringement of the Publishers’ copyrights.” Compl. ¶ 222. That is just a “[t]hreadbare recital[] of” an “element[] of [the] cause of action,” insufficient to state a claim. *Ashcroft*, 556 U.S. at 678. Nor do allegations elsewhere in the Complaint suggest that Microsoft intended, by removing CMI, to further any infringement.

Alleged infringement during training and development. With respect to Microsoft’s alleged conduct in training the GPT models, the Complaint alleges various acts involving “scraping,” “storing,” “processing,” and “reproducing” works to “build [a] training dataset[.]” Compl. ¶¶ 193-94; *see* Compl. ¶¶ 157, 161. It also alleges that “storing, processing, and reproducing the GPT models” themselves constitutes infringement of plaintiffs’ works. Compl. ¶ 195. But it makes no sense to base a CMI claim on alleged infringement during training. The point of CMI, and thus § 1202’s protections, is to “inform *the public* that something is copyrighted.” *Roberts v. BroadwayHD LLC*, 518 F. Supp. 3d 719, 737 (S.D.N.Y. 2021) (emphasis added); *Fashion Nova, LLC v. Blush Mark, Inc.*, No. 22-cv-6127, 2023 WL 4307646, at *5 (C.D. Cal. June 30, 2023) (“[T]he purpose of CMI is to provide *the public* with notice that a work is copyrighted.”) (emphasis added). Training takes place out of public view, and training data itself is not disseminated. That is why the court in *Tremblay* rejected the theory that “knowing removal of CMI from … books during the training process” could violate § 1202. 2024 WL 557720, at *4. The Complaint does not plausibly allege how the training or

development of the models could somehow facilitate Defendants' or anyone else's alleged infringement: The Complaint provides no information about *whom* removal of CMI would conceal infringement *from*, or *how* removal could conceivably make it easier for anyone to train a model using copyrighted works when training takes place outside of public view.

In any event, the Complaint concedes away any claim that Microsoft intends to facilitate infringement during training by acknowledging Microsoft's strong belief that it is "entitled to copy and use any written product" for the purpose of "train[ing] [its] GenAI systems." Compl. ¶ 14. Microsoft bases this belief on (among other things) the doctrine of fair use. As § 107 of the Copyright Act makes plain, "the fair use of a copyrighted work ... *is not an infringement of copyright,*" (emphasis added). And the Complaint does not allege that Microsoft's belief in its rights is ingenuine or objectively unreasonable. A defendant that believes its conduct is lawful cannot also believe it is infringing. Microsoft's good-faith, objectively reasonable belief that it is not infringing thus negates any inference that it knew or should have known it was facilitating or concealing infringement by allegedly removing CMI. *Cf. Mango*, 970 F.3d at 173 (relying on defendant's acknowledgement that "he was required to get permission" to conclude that defendant should have known that removal of CMI would "wrongfully impl[y] ... permission").

Allegedly infringing output. The Complaint also fails to allege scienter based on alleged outputs from a GPT-based product. This is for much the same reason Plaintiffs' theory of contributory infringement fails: The Complaint does not plausibly allege actual user conduct that yields likely infringing outputs in the first place, so it cannot possibly raise the inference that Microsoft somehow intended, knew about, or should have known about such behavior. *See supra* 9. And even if the Complaint did allege user conduct that might yield infringing outputs,

it offers no theory of how the alleged absence of CMI would in any way facilitate that infringement or make it harder to police it than it would be if CMI were preserved.

The Complaint instead describes how copies of Plaintiffs' works *could*, in theory, be output by GPT-based products, Compl. ¶¶ 144-45, 147, and alleges that Defendants have policies that limit or address such outputs, Compl. ¶¶ 153-56. But taking these allegations as true, they make any infringement *less* likely, and demonstrate that Defendants have taken steps to *avoid* any incidental effects of CMI removal—the opposite of what Plaintiffs must allege to establish that Defendants acted with knowledge that removal of CMI would facilitate infringement.

In any event, the only examples the Complaint provides of the GPT-based products outputting Plaintiffs' works are those initiated by Plaintiffs' own contrived prompting, all of which involved Plaintiffs *inputting* or *requesting* content from their own newspapers. If someone already has a particular article with which to prompt ChatGPT, or already has the title of the article to enter as a prompt, the presence or absence of CMI is irrelevant. The user already knows where the article is from. The Complaint does not explain how the absence of CMI in this situation could ever cause the user to behave differently or hinder efforts to thwart infringement.

The Ninth Circuit's decision in *Stevens v. Corelogic, Inc.* is on point. 899 F.3d 666 (9th Cir. 2018). *Stevens* involved the defendant's removal of CMI metadata from photographs as part of technological “[d]ownsampling”—*i.e.*, compression. *Id.* at 671. Plaintiff photographers suggested that this somehow would aid infringement. But they “ha[d] not ... averred that they ha[d] ever used CMI metadata to prevent or detect copyright infringement.” *Id.* at 675. Nor could they establish that the defendant was aware of a third-party “pattern of conduct” or “established modus operandi” giving rise to an inference that the defendant was “aware of the

probable future impact of its actions” with respect to CMI. *Id.* at 674. The plaintiffs therefore could not establish that any removal of CMI was done with an awareness that it would facilitate infringement. *Id.* at 675. Courts have not hesitated to apply *Stevens* at the pleading stage, and doing so is appropriate here.²

B. The Complaint Does Not Plausibly Allege “Removal” Of CMI From Copies Of Identical Works.

Plaintiffs’ CMI claims fail for the further reason that the Complaint does not allege the active “removal” of CMI from an identical copy of a work, as required under § 1202(b). *See Doe I v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, at *8-9 (N.D. Cal. Jan. 22, 2024) (“Section 1202(b) claims require that copies be identical.”) (internal quotation marks omitted). Because it is an anti-piracy statute, § 1202 focuses on removal of CMI *from copies* of works that could potentially be pirated. Several aspects of the text bear this out. First, § 1202 explicitly defines CMI as “information conveyed in connection *with copies* … of a work,” (emphasis added)—that is, in connection with the “material objects” in which the work is “fixed,” 17 U.S.C. § 101—and not, more abstractly, with the underlying work itself. And second, § 1202(b)(1) and (b)(3) use the term “remove,” which connotes some active conduct. To “remove” is “to move by lifting, pushing aside, or taking away or off” or “to get rid of.” Webster’s Third New International Dictionary 1921 (2002).

It makes no sense to speak of *removing* CMI from something that is different than the work with which the CMI was conveyed. One cannot “remove” something where it never existed in the first place. The plain meaning of the noun “copy” is an “imitation,” “transcri[ption],” “reproduc[tion],” or otherwise the product of “duplicat[ion].” Webster’s Third

² See *Tremblay*, 2024 WL 557720, at *4-5; *Andersen*, 2023 WL 7132064, at *10; *Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1174-75 (N.D. Cal. 2019).

New International Dictionary 504 (2002). “Copy” implies sameness, not difference. If Congress had intended that § 1202 cover claims involving altered versions, it could have added words to encompass “alterations,” “abridgments,” “adaptations,” “derivatives,” or any of a number of other words previously used by Congress in the Copyright Act to indicate something other than a copy. That Congress did not do so indicates an intent that the claim address only identical rather than modified versions of a work.

Applying these statutory terms, courts have consistently rejected § 1202(b) claims where a defendant allegedly removed CMI from—or merely failed to transpose CMI to—an excerpt or modification of the original copy with which the CMI was included. Courts have thus rejected DMCA claims relating to the same underlying technology in question here based on the copying of excerpts or modified versions of underlying material. *Doe I*, 2024 WL 235217, at *8-9 (holding in the context of Microsoft AI technology that “no DMCA violation exists unless the works are identical”) (citation omitted). Applying the same principle, courts have rejected claims based on mere “framing” of a photograph in a way that does not include CMI, *Falkner v. Gen. Motors LLC*, 393 F. Supp. 3d 927, 938-39 (C.D. Cal. 2018); excerpting lecture notes and study questions from textbooks without reproducing CMI, *Faulkner Press, L.L.C. v. Class Notes, L.L.C.*, 756 F. Supp. 2d 1352, 1356, 1359 (N.D. Fla. 2010); copying “aspects” of architectural works but “omitting” the plaintiff’s CMI, *Design Basics, LLC v. WK Olson Architects, Inc.*, No. 17-cv-7432, 2019 WL 527535, at *5 (N.D. Ill. Feb. 11, 2019); or incorporating the underlying content from the original copy into some different form or distinct work without CMI, *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (thumbnail versions of images), *rev’d on other grounds*, 336 F.3d 811 (9th Cir. 2003). In short, a CMI-removal claim lies only

with respect to copies of works that are “substantially or entirely reproduced.” *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018).

Similarly, courts have rejected § 1202(b) claims where a defendant “created [a] derivative work” without transposing CMI, because that cannot establish that the defendant “removed [CMI] from [the] original work.” *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-cv-00496, 2015 WL 263556, at *3-4 (D. Haw. Jan. 21, 2015), *aff’d*, 700 F. App’x 674 (9th Cir. 2017). And they have emphasized that § 1202 “does not prohibit merely omitting CMI from an infringing work.” *Dolls Kill, Inc. v. Zoetop Bus. Co.*, No. 22-cv-01463, 2022 WL 16961477, at *3 (C.D. Cal. Aug. 25, 2022); *see A’Lor Int’l, Ltd. v. Tappers Fine Jewelry, Inc.*, No. 12-cv-02215, 2012 WL 12921035, at *10 (C.D. Cal. Aug. 8, 2012) (“omissions” of CMI not actionable). The Complaint fails to identify identical copies from which CMI was removed.

Synthetic search results. With respect to alleged removal of CMI from “synthetic search results” returned by the GPT-based Bing tools, the Complaint vaguely refers to “copies” and “derivatives of the Publishers’ Works,” without explaining what precisely is at issue. Compl. ¶ 219. But elsewhere the Complaint admits what these synthetic search results really are: In “answer [to] user queries,” the synthetic results “may include extensive paraphrases and direct quotes.” Compl. ¶ 69. The mere non-inclusion of CMI with “paraphrases” or “quotes” does not constitute removal of CMI from a substantially identical copy of a work. *Doe I*, 2024 WL 235217, at *8-9.

End-user outputs. Similarly deficient are the Complaint’s allegations with respect to ChatGPT outputs—the ones returned in response to explicit requests for content from specific articles in plaintiffs’ specific newspapers. As the Complaint acknowledges, even these prompts generate only “verbatim *excerpts*” (which, of course, the prompter already knows are from the

specific article in the specific newspaper requested in the prompt). *E.g.*, Compl. ¶¶ 99 (emphasis added), 101, 103, 105, 107, 109, 111; *see* Compl. ¶ 96 (“near-verbatim copies of *significant portions* of the Publishers’ Works”) (emphasis added). Many of Plaintiffs’ prompts specifically *request only excerpts*. Compl. ¶¶ 100, 102, 104, 106 (requesting “the first five paragraphs” of specific articles). The non-inclusion of CMI from an excerpt of a work is nothing like tearing out a book’s copyright page or obscuring the gutter credit on a photograph. The latter acts involve active conduct directed at the CMI on a copy of a work. The functioning of the GPT-based products, as alleged in the Complaint, at most amounts to incidental omission of CMI from a short generated passage of a work. *See supra* 19 (collecting cases).

Training materials. This leaves only the Complaint’s conclusory allegation that “[o]n information and belief, Defendants removed” CMI “in building the training datasets.” Compl. ¶ 218. It is not clear what Plaintiffs are alleging here, because nothing in the Complaint describes the “scraping” or “building” that Plaintiffs allege. This vague allegation fails to identify any actual copy from which CMI is allegedly removed. And in any event, as explained above (at 14-15), the act of removal of CMI during the training process could never give rise to a claim because training takes place out of public view and training data is not disseminated.

The Complaint’s failure to allege removal of CMI from a substantially identical copy further merits dismissal of the § 1202(b)(1) and (b)(3) claims.

III. PLAINTIFFS’ MISAPPROPRIATION CLAIM IS PREEMPTED BY THE COPYRIGHT ACT.

Plaintiffs’ state-law misappropriation claim should be dismissed as preempted by § 301 of the Copyright Act. *See* Compl. ¶¶ 226-32 (Count VI). Section 301 preempts “legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright … and come within the subject matter of copyright.” 17 U.S.C. § 301(a). It thus

operates to bar a claim that sounds in copyright but masquerades as a state law tort in order to avoid copyright law's limits—for example, the fair-use defense that safeguards a wide range of interests by inherently limiting the scope of the copyright monopoly.

Plaintiffs here attempt to avoid preemption by advancing a so-called “hot news” claim, based on the contention that Defendants misappropriate Plaintiffs’ “time-sensitive breaking news.” Compl. ¶ 227. The Complaint alleges that “[t]he Publishers gather information, which often takes the form of time-sensitive breaking news, for their content at a substantial cost to the Publishers.” Compl. ¶ 227. It then alleges that Bing Chat and Browse with Bing can be used for “unauthorized retrieval and dissemination of hot news.” Compl. at 49 (capitalization omitted). In the examples the Complaint provides, Plaintiffs specifically requested either Bing Chat or Browse with Bing to provide a particular article by a Plaintiff that “may not have been included in the LLMs’ training set.” Compl. ¶ 114. The notion seems to be that by combining Bing’s search engine functionality with a GPT-based model, the tool could theoretically retrieve information from recent news articles not in the training set.

These allegations fail to plead the sort of “narrow ‘hot-news’ misappropriation claim [that] survives preemption” under *National Basketball Ass’n v. Motorola, Inc.*, 105 F.3d 841, 852 (2d Cir. 1997). As in *NBA*, Plaintiffs’ claim here implicates the “subject matter” of copyright—the first requirement of the preemption analysis. Although Plaintiffs’ news content may include both “copyrightable” expression and “uncopyrightable events [and] facts,” both of these come “within the ambit of copyright protection” for preemption purposes. 105 F.3d at 848-49. (Were it otherwise, a plaintiff could use state law to evade § 102(b) of the Copyright Act’s carefully drawn limits on protectable subject matter.)

As for the second inquiry—whether the claim “involves extra elements” and therefore asserts rights that are “not the equivalent of exclusive rights under a copyright”—*NBA* explained precisely what a plaintiff would need to allege in order to evade preemption: “(i) the time-sensitive value of factual information, (ii) the free-riding by a defendant, and (iii) the threat to the very existence of the product or service provided by the plaintiff.” 105 F.3d at 850, 853. The Complaint’s attempt to make out these elements is woefully deficient. While it references “time-sensitive” news, the Complaint neither defines that term nor alleges that Defendants’ tools disseminate that time-sensitive news “precisely at the point where the profit is to be reaped.” *Id.* at 851 (internal quotation marks omitted). The Complaint therefore alleges “neither the quantity of copying nor the immediacy of distribution necessary to sustain a ‘hot’ news claim.” *Fin. Info., Inc. v. Moody’s Invs. Serv., Inc.*, 808 F.2d 204, 209 (2d Cir. 1986).

The story told by the Complaint of years of supercomputer development and training of models prior to release is inconsistent with the premise that the models are spitting out today’s news; in fact, it casts considerable doubt on its plausibility. But even if the premise were plausible, the Complaint’s naked assertions that GPT-based tools “directly compete with the Publishers’ content,” Compl. ¶ 228, and “has also reduced the Publishers’ incentive to create such time-sensitive materials,” Compl. ¶ 231, are not. Nothing in the Complaint remotely substantiates these assertions with real-world facts. The Complaint’s assertion that GPT-based tools somehow “siphon[] off existing and potential” subscribers, Compl. ¶ 185, falls flat given the absence of any allegation of any real-world examples of someone using such tools to access Plaintiffs’ works. *See supra* 6-7. The Complaint thus fails to allege the GPT-based tools “would so reduce the incentive to produce” its content “that its existence or quality would be substantially threatened.” *NBA*, 105 F.3d at 845.

Courts applying *NBA* have routinely found hot-news claims preempted at the pleading stage. *E.g., ML Genius Holdings LLC v. Google LLC*, No. 20-3113, 2022 WL 710744, at *5-6 (2d Cir. Mar. 10, 2022); *DBW Partners, LLC v. Mkt. Sec., LLC*, No. 22-cv-1333, 2023 WL 2610498, at *5 (D.D.C. Mar. 23, 2023); *see also Greer v. Fox News Media*, No. 22-1970, 2023 WL 2671796, at *2 (2d Cir. Mar. 29, 2023) (finding preemption as a matter of law). This Court should do so here.

IV. PLAINTIFFS' STATE LAW DILUTION CLAIM IS BARRED BY THE DORMANT COMMERCE CLAUSE.

Plaintiffs' claim for dilution of business reputation under New York's Gen. Bus. Law § 360-1 (Count VIII) must be dismissed because applying state law in the manner Plaintiffs request would be an unconstitutional attempt to directly regulate transactions with no connection to the State of New York.

Under the dormant Commerce Clause of the U.S. Constitution, a state may not "directly regulate[] out-of-state transactions." *Nat'l Pork Producers Council v. Ross*, 598 U.S. 356, 376 n.1 (2023) (emphasis omitted). The Constitution "precludes the application of a state statute" that directly controls "conduct beyond the boundaries of the State," thus guarding against "inconsistent legislation arising from the projection of one state regulatory regime into the jurisdiction of another State." *Healy v. Beer Inst. Inc.*, 491 U.S. 324, 336-37 (1989); *see also Edgar v. MITE Corp.*, 457 U.S. 624, 641 (1982) (declining to enforce an Illinois securities law that "directly regulate[d] transactions which [took] place ... wholly outside the State"). The dormant Commerce Clause applies with special force in the context of regulation of the internet, which "represents an instrument of interstate commerce" that necessarily traverses state lines. *Am. Libraries Ass'n. v. Pataki*, 969 F. Supp. 160, 173 (S.D.N.Y. 1997).

Courts have accordingly held that state laws regulating internet companies violate the dormant Commerce Clause when they directly control what the company may show users on a nationwide basis. In *American Booksellers Foundation v. Dean*, for example, the Second Circuit held that a Vermont statute prohibiting the dissemination of “indecent material” to minors violated the dormant Commerce Clause as well as the First Amendment. 342 F.3d 96, 102-03 (2d Cir. 2003). The Court reasoned that a company providing material on the internet nationwide “must assume that someone from Vermont may also view the material,” which means that out-of-state companies “must comply” with the Vermont statute or face liability under Vermont law. *Id.* The state had therefore “projected its legislation into other States, and directly regulated commerce therein, in violation of the dormant Commerce Clause.” *Id.* at 104 (citation omitted). Other courts adopt similar reasoning. See *ACLU v. Johnson*, 194 F.3d 1149, 1161 (10th Cir. 1999) (New Mexico statute prohibiting internet dissemination of “material harmful to minors” was an “attempt to regulate interstate conduct occurring outside New Mexico’s borders, and is accordingly a per se violation of the Commerce Clause”); *Publius v. Boyer-Vine*, 237 F. Supp. 3d 997, 1025 (E.D. Cal. 2017) (state statute that made it unlawful to post online home addresses and telephone numbers of certain government officials violates the dormant Commerce Clause); *Ctr. for Democracy & Tech. v. Pappert*, 337 F. Supp. 2d 606, 663 (E.D. Pa. 2004) (state statute requiring ISPs to block access to websites displaying child pornography “violates the dormant Commerce Clause”).

The same analysis precludes Plaintiffs’ proposed application of New York Gen. Bus. Law § 360-1 here. Plaintiffs seek to impose liability on Defendants under state law for creating and operating “GPT-based products” on a nationwide basis that allegedly “produce inaccurate content that is falsely attributed” to Defendants. Compl. ¶ 253. And they request an injunction

under state law to preclude Defendants from doing so. Compl. ¶ 254. This is an attempt to control the interactions between users of generative search technology and Microsoft despite their lack of any connection to the state of New York. The “GPT-based products” at issue operate the same way everywhere, by responding to user prompts based on highly sophisticated models trained on a vast “general corpus” of material. Compl. ¶ 76; *see* Compl. ¶¶ 72-78. How the models are trained and how they generate outputs in response to prompts are the same for users in New York as they are for users in any other state. Thus, Plaintiffs’ theory that the New York statute can police the content of ChatGPT’s outputs represents an effort to use one state’s law to directly control nationwide conduct. Count VIII must therefore be dismissed. To the extent Plaintiffs have any trademark-based claim, it must proceed under federal law.

CONCLUSION

The Court should grant the partial motion to dismiss in its entirety.

Dated: New York, New York
June 11, 2024

Orrick, Herrington & Sutcliffe LLP

By: /s/ Annette L. Hurst
Annette L. Hurst (*Pro Hac Vice*)
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
Facsimile: (415) 773-5759
ahurst@orrick.com

Christopher J. Cariello
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
ccariello@orrick.com

Sheryl Koval Garko (*Pro Hac Vice*)
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800

Facsimile: (617) 880-1801
sgarko@orrick.com

Laura Brooks Najemy (*Pro Hac Vice*)
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800
lnajemy@orrick.com

*Attorneys for Defendant
Microsoft Corporation*